Slides & papers

LLMs as Sapience without Sentience and the Problem of Testimonial Knowledge about Other Minds

An Immanent Critique of Simonelli (2026)

Injin Woo & Donggeon Kim (Sungkyunkwan University)

The tension. Two things we already accept pull against each other. We can learn from machines: a hearer who takes the boiling point of water from a language model has gained knowledge. And we can know that a purely text-mediated correspondent — a lifelong pen pal, known through nothing but letters — is someone for whom there is something it is like to write them. If Simonelli's case for sapience without sentience succeeds, we must choose. **Thesis:** granted its own premises, SwS entails that pure LLMs assert and testify — and for that very reason our attribution of phenomenal consciousness to a text-mediated interlocutor ceases to be *knowledge*. Simonelli must restrict his thesis or accept skepticism about other minds.

This handout carries the argument skeleton only. Grounds, textual evidence, and the treatment of the literature are in the paper.

I. The question, the target, the strategy

A. Hinton to LeCun (2023): "The central issue on which we disagree is whether LLMs actually understand what they are saying. You think they definitely don't and I think they probably do." (as cited in Simonelli, 2026, p. 2)

B. **Aim: Simonelli (2026)**
He argues that, given Global Strong Inferentialism (GSI), a pure LLM trained on nothing but linguistic data can possess *every* concept without any conscious awareness or sentience. It can be sapient without being sentient. Call this the SwS thesis.

C. **Our strategy: immanent critique**
The existing critical literature contests the premises, and stalls: the Main Argument is a conditional, and each reply runs it as a modus tollens against a premise Simonelli has already declared himself willing to defend at any intuitive cost. We do the opposite. We grant the premises in their strongest form and ask what follows that he has *not* priced in.

D. **What we do not claim:**

1. Not that hyper-inferentialism is false. Only that endorsing it costs more than Simonelli has priced in.
2. Not that LLMs fail to mean anything, and not that they fail to assert. *Meaningfulness* and *assertoric force* are two distinct properties, and there is a real position that grants the first while denying the second (Borg 2026, §III). We grant both. Two disclaimers follow, one on each side.
   a. Granting assertability is not a *refutation of assertion skepticism*. Borg's road is closed to Simonelli (§III) but open to us, and we simply decline to take it.
   b. Charging SwS with a cost is not a *refutation of assertion optimism* either. Our conclusion is conditional: if assertion skepticism is right, SwS is blocked earlier and our argument never fires. What we contest is only the claim that SwS, even if true, comes cheap.

3. Not a new skeptical scenario. About the traditional problem of other minds we have nothing further to say.

II. Simonelli's puzzle

A. **GSI (Global Strong Inferentialism):**
Mastery of a concept's inferential role is sufficient for possessing it — for *all* concepts, with no exception for perceptual or affective vocabulary. Content words work like logical constants: constituted by substantive inferential rules.

intra-familial: relate a color concept to others in the family
$\text{scarlet}(t) \vdash \text{red}(t)$
$\text{red}(t) \vdash \text{colored}(t)$
extra-familial: relate it to concepts outside the family; what is paradigmatically red
$\text{stop sign}(t) \vdash \text{red}(t)$

B. **Simonelli's radical step:**
He posits no language-entry rules — no bridge principles from perceptual situations to linguistic application ("quasi-inferences"). The standard move adds them; Simonelli refuses. The non-inferential application of "red" is itself captured inferentially, all the way down.

C. **Effability:**
Every part of the conceptual content of every expression is sayable without remainder.

D. **The Main Argument** (Simonelli, 2026, p. 3; numbering ours)

(MA1) Mastery of inferential role is sufficient for possession of all concepts. [GSI]

(MA2) Training on linguistic data is (in principle) sufficient for mastery of inferential role.

(MA3) ∴ So, training on linguistic data is (in principle) sufficient for possession of all concepts.

(MA4) ∴ So, LLMs, trained on nothing but linguistic data, are (in principle) capable of possessing all concepts.

E. **We touch neither (MA1) nor (MA2).**
(MA4), the SwS thesis, then follows. (MA4) is counterintuitive on its face — the concept of red without ever having seen, of pain without ever having felt — and Simonelli accepts it willingly.

F. **Simonelli's definition of sentience: the robust, flexible responsiveness of an animal navigating a perception-action nexus**(Brandom/Gibson)
The cat perceives the table as to-be-jumped-upon. Functional and ecological throughout; nothing about "what it is like" enters the definition (Gibson himself warns that this vocabulary must not be confused with a "world of consciousness"). A pure LLM, neither physically nor virtually embodied, fails this condition outright.

G. **Note$_{\text{fn. 14}}$:**
The whole argument turns on one footnote. Simonelli professes neutrality on the hard problem, but adds: sentience in his defined sense is "necessary for so-called 'phenomenal consciousness' " (2026, p. 13, n. 14). So, on his own view, whatever lacks sentience in his sense lacks phenomenal consciousness.

III. From meaning to assertion to testimony

A. **Two distinct claims:**
- (a) Pure LLMs produce tokens that are semantically significant.
- (b) Pure LLMs carry assertive force.

  - An optimist about (a) can coherently be a skeptic about (b).

B. **Borg's position: the combination is occupied, not stipulated** (Borg, 2026)
  - *Meaning, yes.* LLM outputs inherit type-level meaning by *semantic deference*: they reuse symbols human practice has already made meaningful, in grammatically coherent and contextually apt ways.
  - *Assertion, no.* Assertion presupposes a norm of truth. An LLM confidently produces plausible falsehoods mixed in with facts, which shows the norm governing its output is word-occurrence probability, not truth.
  - "…even though we should count LLM outputs as meaningful we should not count them as assertions." (Borg, 2026, §3)

C. **GSI leaves no room for the distinction:**
If inferential mastery exhausts conceptual understanding, and genuine language use just is participation in the game of giving and asking for reasons, then a pure LLM with that mastery is *thereby* a genuine language user — and genuine language users assert.

D. **Nor can he borrow Borg's ground:**
His own account of theoretical responsibility — answering challenges, clarifying, *retracting under counter-evidence* — is not the profile of a truth-indifferent system. To deny assertion on Borg's ground, he would have to say that a system that takes theoretical responsibility for p nonetheless does not assert p. His framework cannot state that position.

E. **The scope of this critique:**
  - Suspicion: a critique that turns on the meaningfulness/assertability distinction would seem to apply to every theorist friendly to pure-LLM assertability — so it does not distinctively target Simonelli.
  - Reply: what we are pressing is the weight of the semantics he commits to. Under GSI, once a producer of meaningful tokens counts as a genuine language agent, it cannot fail to assert: on his semantics, meaning-expressibility entails assertability.
  - That puts him at a disadvantage in the semantic debate. Unlike most semantic optimists, he can keep semantic optimism only by committing, always, to a theoretically burdensome pragmatic thesis as well.
  - So the target is not assertability-optimists in general, but only theories carrying an indefeasible commitment linking the semantic thesis to the pragmatic one. Simonelli is its paradigm.

F. **The consequence chain:**
Assert ⇒ testify ⇒ we can acquire epistemic justification from a non-human testifier. Nothing objectionable in general — which is exactly what makes the consequence hard to contain. The problem is located in one place: testimonially based knowledge of *other minds*.

Neither step is neutral common ground; both hold within the granted framework. On Simonelli's own account, the core function of assertion is to entitle others to make that assertion themselves — the structure of testimonial inheritance of entitlement (2026, p. 25) — and he says that calling ChatGPT a knower is taking it to be reliable testimonially in the game of giving and asking for reasons (p. 27). Assurance-theoretic views of testimony may resist the second step, but that resistance blocks SwS earlier, in the same place as assertion skepticism (§I, on the

conditional scope of our conclusion).

G. **Why the target has to be Simonelli:**
What the argument below needs is two halves at once: full discursive competence and the absence of sentience. SwS is the only position that certifies both. Assertion skepticism blocks the first half; an embodiment requirement blocks the second. The mere existence of fluent LLMs does not get the argument off the ground.

IV. The argument from other minds

A. **The test case: a lifelong pen pal**

- Forty years of letters and nothing else — no face, no voice, never a shared room. Ordinarily we say you *know* there is a mind at the other end: not merely that your correspondent is clever, but that there is something it is like for her to sit down and write to you.
- The pen pal stands in for every interlocutor to whom text is the only access: the faceless remote collaborator, the online-only colleague — and the pure LLM itself.
- Origin: the scenario is our construction; its epistemology is not. Gomes (2015) argues that testimony can be a *basic* source of knowledge of another's mental life, with attributions of affective states — tiredness, boredom — as his central cases. The pen pal pushes that thesis to its limiting case.

B. **The Argument from Other Minds:**

(OM1) A pure LLM is a fully competent discursive performer and lacks phenomenal consciousness. [Simonelli's thesis, granted]

(OM2) If (OM1), then discursive performance is not *by itself* evidence that its producer is phenomenally conscious.

(OM3) Prior to SwS the background generalization held, so text-mediated attributions of phenomenal consciousness could be knowledge.

(OM4) SwS falsifies that generalization while leaving the attributor's evidence untouched, so a true attribution formed now is true by luck, hence not *knowledge*.

(OM5) No non-discursive evidence available within Simonelli's own framework restores it.

(OMC) ∴ Simonelli must either restrict SwS or accept skepticism about *knowledge* of text-mediated other minds. [from OM3–OM5]

C. **Defending the premises**

- (OM1) The actuality claim is his own: he names ChatGPT as bearer of the siphoned notion of theoretical responsibility (2026, p. 27), and hopes present systems will *stay* sapient without becoming sentient (p. 31) — which cannot be said without a verdict about the actual case. That such a system lacks phenomenal consciousness follows from his sentience definition together with fn. 14 (§II). That footnote does double duty: it yields the premise here and closes the first exit below, since a necessary condition settles absence but never presence.
- (OM3) The generalization: producing discursive performance of this sophistication is reliable evidence that the producer is sentient. The claim is restricted

to text-mediated attribution. And what the premise asserts is not that such attributions were justified but that they *could be knowledge*: without that baseline there is nothing for (OM4) to report as lost.

- (OM4) The generalization goes false or nearly so while the evidence does not change: reading the same prose from the same pen pal, her belief could now too easily have been false.
- (OM5) Discursive restoration is already closed by (OM2) and (OM4). What remains are non-discursive restorations, and why they come to just three is defended below.

D. **The severance is by design:**

- On Simonelli's own account, full discursive performance is constituted by two things only: inferential mastery (§II) and the capacity to bear theoretical responsibility.
- "We can siphon out the purely theoretical aspects of the responsibility that one takes in making an assertion, and we are left with an intelligible notion of responsibility… that is in principle attributable to LLMs like ChatGPT" (2026, p. 27).
- Neither constituent refers to sentience. That is not a gap he overlooked; it is the thesis.
- What (OM2) denies is therefore only the constitutively licensed inference; whatever contingent evidential force performance retains runs through the background generalization in (OM3).

E. **The pen pal, and the background generalization:**

- Where no channel but text exists, discursive performance is not one evidential source among several; it is the only one.
- That performance carries evidential weight only because the background generalization holds: *sophisticated discursive performance is reliable evidence of sentience*.
- No adjudication among the accounts is needed. Any account on which text-mediated phenomenal attribution can be knowledge either requires something phenomenal on the speaker's side or does not. If it does, what SwS removes from a pure LLM is exactly that. If it does not, then on that account text was never evidence of the speaker's phenomenal consciousness in the first place, and our conclusion already holds. The division is what carries the argument, not the number of accounts. The three below are the ones that actually occupy the first branch.
  - If testimony is a basic source of knowledge of minds (Gomes, 2015), what lets testimony *originate* knowledge is the speaker's own non-testimonial access to her states. A pure LLM has no such states to access.
  - If the work is done by the hearer's grasp of expressive behavior (Parrott, 2026), what makes that behavior evidentially good is its falling within an interval in which the speaker instantiates the state expressed. A pure LLM has no state that could make an interval that interval.
  - If what justifies the hearer is, externalistically, the speaker's antecedent mental state (Munro, 2022), a pure LLM supplies none.
- Attributions to animals and pre-linguistic infants are untouched: they rest on a different evidential base, embodied behavior, and shared physiology and evo-

lutionary lineage license the inference from function to phenomenology independently.

F. **The generalization never worked alone:**
   - Its reliability was parasitic on the composition of the producer population.
   - The only producers of discourse at this level were human beings, and of human beings we independently take there to be something it is like to be them. That conjunction was the support.
   - SwS falsifies the first fact and leaves the second untouched. Remove one support and the reliability goes with it.

G. **Why this is a knowledge failure, and new:**
   - What does not change: the attributor's evidence. She reads the same prose from the same pen pal.
   - What changes: the environment the evidence sits in. A true belief formed in an environment saturated with indistinguishable non-sentient producers is true by luck (Goldman, 1976; cf. Pritchard, 2005).
   - The traditional problem of other minds never touched the generalization: it traded on mere possibility, while SwS asserts, on independent theoretical grounds, that the profile is *actual*.
   - So what Simonelli supplies is an actualized defeater, not one more speculative skeptical possibility.

H. **(OM5): three exits, no fourth**
   The three are not a list we chose; they are what is left after the divisions. A restoration either offers evidence or denies that evidence is needed. If it offers evidence, that evidence is about the text or about something else. The text is already closed by (OM2) and (OM4). If it is about something else, the fact concerns either what the producer now is or how it came to be that way, and between synchronic and diachronic there is no third: the former is the interlocutor's *body*, the latter the *etiology* of the output. A fact about the composition of the producer population is not a fourth place but the very generalization (OM4) has destroyed. If evidence is denied to be needed, the route is unmediated *acquaintance* or a default entitlement given without evidence — and what closes the entitlement route is (OM4), not (OM5), since environmental luck is indifferent to the kind of warrant. Third-party testimony and institutional certification reduce to what the third party knows, and that knowledge falls under the same divisions again. Three places are therefore left to be closed separately. The literature that actually occupies each is named below.

   1. *Body.* **The exit:** look past the text to what the system is — sensory grounding (Chalmers, 2023), a biological substrate (Seth, 2025), and, in the refined version, deep computational markers such as a global workspace (Birch, 2024, Proposal 24).
      - **Blocked in the text-mediated case by stipulation.** The pen pal is known through letters and nothing else; a pure LLM is neither physically nor virtually embodied.
      - **Idle even where the body is available.** Sentience in his sense is only *necessary* for phenomenal consciousness, so confirming the architecture confirms at most that a system *might* be conscious. His own hedge concedes this: a perception-action architecture "could in principle" be sentient (2026, p. 16).

  - **Computational markers inherit the limit.** They are not a fourth place to look but a refinement of the first, since a system's computational organization is a fact about what the system is. However good interpretability gets, what it delivers is a functional fact, and functional facts secure only the necessary condition.

2. *Etiology.* **The exit:** appeal to the causal history of the output — production goals, training regime, computational structure — and let sentience attribution answer to that rather than to performance. What opens this place is Birch's gaming problem: any marker may have been gamed, so we must ask how it came about (Birch, 2024, Proposal 23). Birch's own answer moves back to the first place, deep computational markers (Proposal 24), which is already a sign that this exit does not stand on its own. Note that this exit is stronger than it first looks: Simonelli can consistently hold that sapience is constituted by performance while sentience is an empirical matter, so the charge that etiology "conflicts with his performance-based criterion" does not by itself close it.
   - **It relocates the verdict without delivering one.** Taking the exit concedes that sentience attribution answers to facts outside discursive practice; the question is what those facts then say.
   - **They say nothing decidable.** The systems at issue are optimized to produce text that reads as sincere, so the right causal history is no more verifiable than the wrong one — which the precautionary literature concedes even while hoping to overcome it (Birch, 2024, pp. 320–321).
   - **So the exit is open but empty.** In text-mediated cases causal history is not evidence we are in principle denied; it is evidence we do not have.
3. *Acquaintance.* **The exit:** hold that our grip on another's consciousness rests on something that discursive performance does not exhaust — a direct relation, or a residue of content that cannot be said. Chalmers occupies it directly: on his two-tiered inferentialism the first tier of contents derives from *acquaintance with experience*, and in a pure thinker that tier is mostly absent (Chalmers, 2023). On the other-minds side the nearest occupants are expressive knowledge that is neither evidential nor perceptual (Gomes, 2019) and expression disjunctivism (Parrott, 2026).
   - **Foreclosed by his own effability thesis** (§II): every part of the conceptual content of every expression is sayable without remainder.
   - **Taking it costs him the framework.** An ineffable residue doing evidential work that discursive performance cannot do is precisely the quasi-inference GSI was built to avoid; readmitting it here readmits it everywhere.

I. **Anticipated reply, and why it fails:**
   - The reply: "I never said our confidence rests on anything ineffable, and I owe no account of what does ground it; I am neutral on the hard problem."
   - Why it fails: the argument does not ask him what *does* ground other-minds attribution. It asks him to notice what his own theory entails *cannot*.
   - Verdict: declining to say what does ground it is not neutrality once one's theory has said, in detail, what does not.

V. Two costly options

A. **Restrict SwS**
   - The move: exempt phenomenal vocabulary — for "pain," "sees red," inferential mastery is not enough; acquaintance or a language-entry rule is also required.
   - The cost: reintroduces the quasi-inferential rules GSI was built without, with no principled place to stop. The project collapses into a far more modest thesis.

B. **Accept skepticism**
   - The move: give up knowledge of other minds' phenomenal consciousness in text-mediated cases.
   - The cost: collides with his own final section — we must not build sentient systems — a norm his own necessary-condition admission (§IV) leaves him unable to act on for the very architectures it targets.

C. **Why the second cost is the heavier one**
   - What the norm targets is phenomenal consciousness, not functional sentience. Of the two grounds he gives, the safety one holds for functional goal-pursuit alone, and we grant it. The ethical one — that we would incur obligations to these beings and almost certainly fail them — requires that they be moral patients, and what makes a moral patient is not a perception-action nexus but well-being of which one is aware. His own next sentence after the functional characterization says just this (2026, pp. 29–30).
   - For a pure LLM the norm has no work to do: architecture alone settles the verdict, so there is nothing to misidentify.
   - The norm does its work on the perception-action architectures he allows "could in principle" be sentient (2026, p. 16). There, confirming the architecture confirms the necessary condition and leaves the phenomenal fact the norm asks about unconfirmed.
   - So the prescription is idle not for want of a target, but because for the one class of target that matters he has no way to tell.
   - The first option carries no such cost. Require acquaintance for phenomenal vocabulary and mastery of that vocabulary marks consciousness again. Only the option that keeps generality recovers the mark nowhere.

VI. Conclusion

The move that lets us gain knowledge from machines costs us knowledge of persons.

A coda: this work also brings out that earlier philosophy of language implicitly relied on the contingent fact that genuine language agents were humans like us (cf. Cappelen & Dever, 2021). If non-human language agents are possible, the semantic framework needs to be finer-grained — and the meaningfulness/assertability distinction is one way of making it so.

References

Birch, J. (2024). *The edge of sentience: Risk and precaution in humans, other animals, and AI*. Oxford University Press.

Cappelen, H., & Dever, J. (2021). *Making AI intelligible: Philosophical foundations*. Oxford University Press.
Chalmers, D. J. (2023). Does thought require sensory grounding? From pure thinkers to large language models. *Proceedings and Addresses of the APA, 97*, 22–45.
Borg, E. (2026). LLMs, Turing tests and Chinese rooms: The prospects for meaning in large language models. *Inquiry, 69*(6), 2807–2837.
Brandom, R. (1994). *Making it explicit: Reasoning, representing, and discursive commitment*. Harvard University Press.
Gibson, J. J. (1979). The theory of affordances. In *The ecological approach to visual perception* (chap. 8).
Goldman, A. I. (1976). Discrimination and perceptual knowledge. *The Journal of Philosophy, 73*(20), 771–791.
Gomes, A. (2015). Testimony and other minds. *Erkenntnis, 80*(1), 173–183.
Gomes, A. (2019). Perception, evidence, and our expressive knowledge of others' minds. In A. Avramides & M. Parrott (Eds.), *Knowing other minds* (pp. 148–172). Oxford University Press.
Parrott, M. (2026). Expression, testimony, and other minds. *Philosophical Studies, 183*(1), 287–305.
Pritchard, D. (2005). *Epistemic luck*. Oxford University Press.
Simonelli, R. (2023). How to be a hyper-inferentialist. *Synthese, 202*(5). https://doi.org/10.1007/s11229-023-04382-1
Seth, A. K. (2025). Conscious artificial intelligence and biological naturalism. *Behavioral and Brain Sciences* (forthcoming).
Simonelli, R. (2026). Sapience without sentience: An inferentialist approach to LLMs. *Asian Journal of Philosophy, 5*(1). https://doi.org/10.1007/s44204-026-00400-4